

Journal Pre-proof

Global Food-source Identifier (GFI): Collaborative virtual research environment and shared data catalogue for the foodborne outbreak investigation international community

Ana Sofia Ribeiro Duarte, Cecilie Liv Nielsen, Leonardo Candela, Lars Valentin, Frank M. Aarestrup, Håkan Vigre

PII: S0956-7135(20)30539-9

DOI: <https://doi.org/10.1016/j.foodcont.2020.107623>

Reference: JFCO 107623

To appear in: *Food Control*

Received Date: 9 July 2020

Revised Date: 4 September 2020

Accepted Date: 8 September 2020

Please cite this article as: Ribeiro Duarte A.S., Nielsen C.L., Candela L., Valentin L., Aarestrup F.M. & Vigre Hå., Global Food-source Identifier (GFI): Collaborative virtual research environment and shared data catalogue for the foodborne outbreak investigation international community, *Food Control* (2020), doi: <https://doi.org/10.1016/j.foodcont.2020.107623>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.



Author statement:

Ana Sofia Ribeiro Duarte: Conceptualization, Project administration, Supervision, Formal analysis, Visualization, Writing - original draft. **Cecilie Liv Nielsen:** Investigation, Data curation, Conceptualization, Formal analysis, Writing - original draft. **Leonardo Candela:** Software, Writing - review & editing. **Lars Valentin:** Software, Writing - review & editing. **Frank M. Aarestrup:** Funding acquisition, Conceptualization, Writing - review & editing. **Håkan Vigre:** Project administration, Conceptualization, Writing - review & editing.

Journal Pre-proof

1 Global Food-source Identifier (GFI): collaborative virtual research environment and shared data catalogue for
2 the foodborne outbreak investigation international community

3

4 Ana Sofia Ribeiro Duarte^{1*}, Cecilie Liv Nielsen^{1,4}, Leonardo Candela², Lars Valentin³, Frank M. Aarestrup¹,
5 Håkan Vigre¹

6

7

8 ¹Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark.
9 Kemitorvet 204, DK 2800 Kongens Lyngby, Denmark

10 ²Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", National Research Council of Italy, Via G.
11 Moruzzi, 1 - 56124 Pisa, Italy

12 ³ Unit Food Hygiene and Technologies, Supply Chains, Food Defense, Department Biological Safety,
13 German Federal Institute for Risk Assessment (BfR), Max-Dohrn-Strasse 8-10, 10589, Berlin, Germany

14 ⁴Department of Animal Science – ANIS Welfare, Aarhus University. Blichers Alle 20, DK 8830 Tjele,
15 Denmark.

16

17 *corresponding author - contacts: asrduarte@gmail.com; Kemitorvet 204, DK 2800 Kongens Lyngby,
18 Denmark

19

20

21

22

23

24

25

26 Competing interests:

27 The authors have no competing interests to declare.

28

29 Author contribution:

30 **Ana Sofia Ribeiro Duarte:** Conceptualization, Project administration, Supervision, Formal analysis,
31 Visualization, Writing - original draft. **Cecilie Liv Nielsen:** Investigation, Data curation, Conceptualization,
32 Formal analysis, Writing - original draft. **Leonardo Candela:** Software, Writing - review & editing. **Lars**
33 **Valentin:** Software, Writing - review & editing. **Frank M. Aarestrup:** Funding acquisition, Conceptualization,
34 Writing - review & editing. **Håkan Vigre:** Project administration, Conceptualization, Writing - review & editing.

35 Abstract

36 The source of a foodborne disease outbreak (FBO) is often difficult to identify, especially in the early phase
37 where interventions would be most efficient. In addition, data on FBOs are mostly scattered in different
38 formats either in national databases and reports or within pathogen-specific or regional reporting networks,
39 both of which are often only accessible to a selected number of individuals.

40 Here, we developed an international, open, shared and searchable data catalogue of past FBOs – the Global
41 Food-source Identifier (GFI). GFI was developed with two objectives: a) to create a collaborative online
42 community of FBO investigators, encouraging the international sharing of data in a harmonized, detailed and
43 comparable manner and b) to support foodborne outbreak investigation worldwide by providing access to
44 detailed records of past outbreaks, which can convey valuable insight into potential ‘risk foods’ of a detected
45 pathogen. GFI is hosted within a Virtual Research Environment (VRE), which offers additional features to
46 facilitate the collaboration between the outbreak investigators. These features allow document exchange,
47 communication and data visualization and analysis between the VRE members.

48 Based on scientific literature on foodborne outbreaks and discussions within a working group, we selected a
49 total of 46 attributes characterising the outbreak records to be included in the catalogue, aggregated under
50 the four overarching categories *causative agent*, *epidata*, *food source* and *report details*. Detailed
51 descriptions of the attributes in the catalogue and instructions for harmonized data reporting are available on
52 a wiki page in the VRE. At the time of writing and public launch of GFI, the data catalogue was populated
53 with records of 102 FBOs occurred in Denmark over a period of 12 years (2005-2016) and covering the most
54 frequent pathogens and a broad range of typing methods.

55 The VRE features that enable data analysis, document sharing and communication between members were
56 applied for the graphical representation of the records available in GFI, and for the sharing of results and
57 script files within the VRE. The descriptive analysis included the relationship between the most frequent
58 causative agents and outbreak food sources. Such results can support a risk-based food sampling strategy
59 in the very beginning of a foodborne outbreak investigation.

60 The Global Food-source Identifier is a data catalogue specifically designed to host an international collection
61 of FBO records reported in a detailed and harmonized manner. It is implemented in a virtual research
62 environment that offers key features to facilitate and enhance the global collaboration and data sharing
63 among FBO investigators. Once in active use by the international food safety community, we envisage that
64 GFI will contribute to the success of FBO investigations worldwide.

65 Keywords

66 Foodborne outbreak investigation; Microbial typing; Food source; Data sharing; Harmonization; Virtual
67 research environment

68

69

70

71

72

73

74

75

76

77

78 1. Introduction

79

80 In 2019 the United Nations World Food Safety Day was celebrated for the first time, with the aim to
81 strengthen the global efforts to ensure food safety. Foodborne infectious diseases have been estimated to
82 affect 550 million persons and cause 230,000 deaths globally in 2010, with 43% of the disease burden from
83 contaminated food occurring among children under 5 years of age (WHO, 2015). These estimates are
84 considered conservative due to considerable underreporting, and foodborne outbreaks (FBO) are
85 responsible for a large part of the estimated burden. In addition to these direct consequences FBO can
86 cause major losses due to restrictions on trade and sale of specific food products.

87 FBO are caused by many different agents contaminating a broad range of foods (EFSA and ECDC, 2015).
88 The task of investigating an outbreak is complicated among other factors due to the difficulty to quickly
89 identify and recover the contaminated food. In Europe, the source of 7% of strong-evidence FBO and 73% of
90 weak-evidence FBO was either reported as 'unknown' or 'other foods' (EFSA and ECDC, 2015). Similarly, in
91 the US, 46% of reported outbreaks had an unidentified food vehicle (Richardson et al., 2017). Traceback
92 studies not only allow resolving an outbreak investigation and containing disease spread, they are also
93 important to support food safety policy evidence-based decisions (Gormley et al., 2011). Rapid, and correct
94 identification of the source is absolutely essential, both to limit human illness and to avoid unnecessary
95 restrictions on falsely implicated food products. Further, rapid identification of the most likely source for
96 sampling and subsequent conventional or molecular analysis is the first and essential step in any outbreak
97 investigation.

98 Characterisation of foodborne isolates differs between countries and laboratories, and may include different
99 phenotyping or molecular typing methods. Different protocols for various typing methods exist and some
100 typing results can be retrieved both from conventional protocols and from whole genome sequencing
101 (Joensen et al., 2015). Nomenclature of outbreak strains can also differ between countries and regions
102 (Peters et al., 2017). However, with an increasing evidence for the transboundary nature of foodborne
103 pathogens there is an increased need for international collaboration and data sharing for surveillance and
104 outbreak investigation. Indeed, one of the strategic directions of WHO's plan to lower the burden of infectious
105 foodborne disease is to improve international collaboration and enhance communication, which are
106 recognized major challenges when food supply chains cross multiple national borders (WHO, 2015). In order
107 to address these challenges at a truly global scale and in an all-inclusive manner, solutions that enhance
108 collaboration, communication and data exchange need to encompass a wide range of typing methods and
109 epidemiological data, allow reporting of the typing protocols used, and adopt standardised nomenclature, in
110 order to ensure transparency and comparability of results.

111 Here we addressed the challenge of exchanging typing results and epidemiological data from FBO
112 investigations, in a collaborative, interactive and harmonised way, internationally. We created a free and
113 open access, shared and searchable data catalogue of foodborne outbreaks, where each record includes
114 detailed information on the type of causative agent, the identified food source and epidemiological
115 information. The catalogue is integrated in a virtual research environment (VRE) named Global Food-source
116 Identifier (GFI)¹. A VRE is an open, alive, online system helping people and organisations to easily share
117 information and results, when research and outbreak investigations are carried out in teams which span
118 across institutions and countries. Features within a VRE usually include collaboration support (Web forums
119 and wikis), document hosting, data analysis and data visualization facilities (Candela et al., 2013).

120 GFI is intended to host an online international community involved in outbreak investigation, who can
121 communicate and exchange no confidential data in a harmonized, comparable manner. The main aim of GFI
122 is to assist that international community during outbreak investigation.

123

¹ Global Food-source Identifier Virtual Research Environment dashboard <https://aginfra.d4science.org/web/foodborneoutbreak>

124 To the best of our knowledge, an international networking platform which is open, allows data sharing in a
125 transparent way and includes detailed information on a comprehensive span of pathogen typing results of
126 such a wide range of phenotypic and genotypic typing methods, on food source and on epidemiology, is not
127 available at the moment. A similar concept has been implemented in the past with the Pathogen-annotated
128 Tracking Resource Network (PATRN) (Gopinath et al., 2013), developed in the US by the Center for Food
129 Safety and Applied Nutrition (CFSA) and Food and Drug Administration (FDA). However, to our best
130 knowledge, this resource is no longer available. Several countries have national databases containing
131 information on foodborne, waterborne or infectious disease outbreaks. (Anon., 2017; Gormley et al., 2011;
132 Guzman-Herrador et al., 2016; Krause et al., 2007; Richardson et al., 2017). These databases are structured
133 in different ways and are not interconnected. Communication between countries relies therefore many times
134 on the individual networks of the people involved in national outbreak investigation. There are also several
135 existing international databases that do not match the characteristics of GFI, such as the Foodborne Illness
136 Outbreak Database (Marler Clark, 2018), The European Surveillance System (TESSy) (ECDC, 2019),
137 NoroNet (RIVM, 2018), PulseNet International (Nadon et al., 2017), The Global Microbial Identifier (Wielinga
138 et al., 2017) or GenomeTrakr (Timme et al., 2019). Most of these data collections are targeted at a specific
139 pathogen or typing method, and many are only fully accessible to a limited number of persons or accessible
140 to a limited extent to the public.

141 Here we describe the development of GFI, including i) the data catalogue that includes multiple pathogens
142 and a comprehensive set of typing methods and protocols, and ii) the VRE - a living environment that hosts
143 the data catalogue, and many additional features that facilitate the collaboration between the GFI community
144 members.

145

146

147 2. Methods

148 2.1 Data catalogue of foodborne outbreaks

149 2.1.1 Selection of variables

150 The outbreak record format supported by the catalogue has been developed as parsimoniously as possible,
151 including only variables considered relevant for foodborne outbreak investigation and those generally
152 present in reports of FBO (table 1). The selection was based on a review of scientific literature, advice from
153 two experts in food- and waterborne outbreak investigation from the Danish Statens Serum Institut (SSI) and
154 discussions within a working group, which included three experts on epidemiological investigation of
155 foodborne disease from the National Food Institute (Technical University of Denmark), a food microbiologist
156 from SSI, and a computer scientist expert on virtual research environments development and catalogue
157 exploitation from the Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (National Research
158 Council of Italy).

159

160 2.1.2 Population of the catalogue

161 The catalogue was initially populated with outbreak data described in the Annual Report on Zoonoses in
162 Denmark in the sections of 'Food- and Waterborne outbreaks' from 2013 to 2016, 'Outbreaks of special
163 interest' from 2007 to 2012 and 'Outbreaks of zoonotic gastrointestinal infections' from 2005 to 2006 (Anon.,
164 2006, 2012, 2013, 2017). Outbreaks considered neither of food- nor of waterborne origin were excluded. The
165 intention was to populate the data catalogue with a minimum of 100 outbreaks covering a period of
166 approximately 10 years, the most frequent foodborne pathogens and a broad range of typing methods. This
167 helped to complete as much as possible the list of typing methods and protocols used for different
168 pathogens. Besides the information retrieved from the Danish zoonoses reports, additional data were
169 extracted from the national database of foodborne outbreaks (Fødevareudbrudsdatabase (FUD)) and
170 whenever needed, from scientific literature describing the outbreaks in more detail. For outbreaks reported
171 both in the zoonoses report and in FUD, the data from FUD was preferred if discrepancy between the two
172 sources existed. Scientific literature used included references mentioned in the zoonoses reports and a
173 selection of additional references from a list provided by one of the outbreak investigation experts (Steen
174 Ethelberg, *personal communication*).

175

176 2.2 Virtual research environment for outbreak investigation

177 2.2.1 GFI features

178 The GFI VRE was created to support the establishment of an international collaboration environment for
179 foodborne outbreak investigation, including a data sharing opportunity as well as features that allow
180 document exchange, communication and data visualization and analysis.

181 The VRE is hosted by the AGINFRA Gateway (Assante et al., 2019a), a science gateway developed by
182 relying on the D4Science infrastructure (Assante et al., 2019c) and the gCube technology (Assante et al.,
183 2019b). This VRE realizes a web-based working environment (figure 1a) providing its users with: (a) a
184 *shared workspace*, for storing, organising and sharing any version of an artefact, including dataset and
185 model implementation (figure 1c); (b) a *social networking area* enabling collaborative and open discussions
186 on any topic and disseminating information of interest for the community, e.g. the availability of a new
187 outbreak record (figure 1e); (c) the *catalogue* recording the outbreak records worth being shared thus to
188 make it possible for others to be informed and make use of these assets (figure 1b); (d) an RStudio-based
189 development environment for R enabling to perform statistical computing tasks in the cloud. The
190 environment provides its users with a powerful integrated development environment (IDE) including a
191 console, a source code editor that supports direct code execution, as well as tools for plotting, history,
192 debugging and workspace management (figure 1d); (e) a Jupyter-based notebook environment for

193 documenting and recording analytics processes (figure 1d). Every notebook is a rich document that contains
194 live code, equations, visualizations and narrative text aiming at capturing a research activity.

195 2.2.2 Collaborative data analysis

196 The full content of the data catalogue can be extracted in the format of a .xlsx or the proprietary KNIME table
197 file via the execution of a KNIME workflow integrated into the VRE². The workflow was developed with the
198 software KNIME (version 3.8) (Berthold et al., 2007) and is offered by the *as-a-Service* delivery model. When
199 extracting the data, the user's personal token is automatically provided in the workflow. After execution of the
200 workflow, the user can proceed with analysis of the extracted data within the VRE, using RStudio or Jupyter
201 notebook. Extracted data can be found within the VRE environment³. An additional output file is the KNIME
202 workflow in its executed status, which the user can open in KNIME (version 3.7 and above) to check
203 individual nodes, if wished. Note that sharing the executed workflow file with other users will imply sharing
204 the personal user token, which is not advisable.

205 Visualization plots and descriptive analysis of the present catalogue data were performed using R by
206 RStudio (R version 3.4.4) (R Core Team, 2018; RStudio Team, 2016) within GFI, and the results were both
207 stored in GFI's workspace and shared in the social networking area, to illustrate GFI's potential to support
208 collaborative outbreak investigation. The R script is provided in a R notebook available to users in the shared
209 workspace.

210

211 3. Results

212

213 The catalogue records are characterised by 46 attributes divided under four overarching categories: (i)
214 *causative agent*, (ii) *epidata*, (iii) *food source* and (iv) *report details* (table 1). Additionally, each record is
215 automatically associated to tags for outbreak year, causative agent, country and food source (type, subtype
216 and variety), and a unique record URL is generated. Management info is also added by default, including
217 author, maintainer, version, state, creation date and date of last update. Users are encouraged to attach
218 complementary data and/or resources to the records, including e.g. scientific publications where the
219 outbreak was described, or a related national/international outbreak report. Record visibility may be set as
220 'restricted' (only VRE members will see the full record) or 'public'. GFI members have access to all catalogue
221 records, while external users only have access to public records and the title of restricted records.

222 A detailed description of the catalogue record' attributes with pre-defined choice- or syntax- options is
223 presented in Appendix A. For the category *causative agent* the following conditions applied: (i) fields were
224 included in order to allow a detailed report of typing results; (ii) for each typing method, a list of the most
225 common protocols is provided in a dropdown menu for the user to choose from. The list of protocols was
226 built based on those used in a Danish setting and it can be reviewed and updates requested by the users as
227 different protocols are applied; (iii) whenever applicable, a specific nomenclature format was selected as the
228 GFI standard for reporting of a given typing method, in order to ensure data harmonization and data
229 comparability. For Pulsed-field Gel Electrophoresis (PFGE) and whole genome sequencing (WGS) the user
230 can report whether the outbreak isolate was PFGE typed and with which protocol, and whether it was whole
231 genome sequenced. A PFGE profile or WGS sequence can be easily accessed by users by either contacting
232 the member who reported the outbreak or by consulting the data and resources attached to the record.

233 Food source related attributes were categorized based on a food categorization scheme from the
234 Interagency Food Safety Analytics Collaboration (IFSAC). The scheme is a five-level hierarchy and its
235 categories take into account the implication of different processing, preparation and consumption practices
236 on the microbiological environment of food (Richardson et al., 2017). A detailed description of the food

² KNIME workflow available at: <https://aginfra.d4science.org/group/foodborneoutbreak/data-miner>

³ Data extracted from the catalogue and the R notebook can be accessed and downloaded at:

<https://aginfra.d4science.org/group/aginfra-gateway/workspace?itemid=8b1cf941-9571-4c12-80e7-42c525472f03>

237 categorization scheme is presented in Appendix A16. When reporting data to the catalogue, food sources
238 must be assigned to the most specific possible food category.

239 Tables 2-4 summarize the outbreaks entered in the catalogue at the time of its public launch. The catalogue
240 included at the time of writing 102 FBO with 166 outbreak stains, which originated 13,466 disease cases
241 spanning over 12 years (table 2). *Salmonella enterica* subsp. *enterica*, *Norovirus* and Enterotoxigenic
242 *Escherichia coli* (ETEC) were the most common causative agents. Reported typing results across all
243 causative agents included serotype, PFGE type, phage type, multiple-locus variable number tandem repeat
244 analysis (MLVA) profile, multilocus sequence type (MLST), WGS, antimicrobial resistance (AMR) profile,
245 virulence profile and genogroup (table 3).

246 In table 4 the distribution of food sources causing the reported outbreaks is presented. The major food type
247 'meat-poultry' was reported in 28% (29/102) of the outbreaks, which were mainly caused by beef, 28%
248 (8/29), and pork, 55% (16/29). Only 10% (3/29) of those outbreaks were caused by poultry. The major food
249 type 'produce' was reported in 37% (38/102) of the outbreaks. Of those 24% (9/38) were caused by fruits
250 and 76% (29/38) by vegetables.

251 The R studio notebook that is available to all VRE members in GFI's shared workspace was used to illustrate
252 the visualization of typing methods, outbreak settings, epidemiology and food sources associated with the
253 most common causative agents (*Salmonella* Enteritidis, *Salmonella* Typhimurium (including monophasic),
254 *Escherichia coli* (ETEC and Shiga toxin/verocytotoxin-producing (VTEC/STEC)), *Norovirus* and
255 *Campylobacter jejuni*) (Figures 2-6). It contains explanatory comments that clarify the different chunks of
256 code, so it can be used by members who are non-proficient in R language. It can be executed in its present
257 version on existing or future datasets extracted from the catalogue data, and it can be adapted by any VRE
258 member to different analysis needs.

259 Figure 2 shows a clear increasing trend in whole genome sequencing of *Salmonella enterica* strains after
260 2012, with all strains being sequenced from 2014 onwards. Despite this trend, all strains continued to be
261 MLVA-profiled. Figure 3 shows the distribution of outbreak settings and major food sources for each
262 causative agent. Compared to other agents, outbreaks of the three most common *Salmonella* serotypes
263 were more often associated to disseminated cases (figure 3a). It is also obvious that for some causative
264 agents there has been a predominant major food source of outbreaks – 'meat-poultry' for *Salmonella*
265 Typhimurium, 'eggs' for *Salmonella* Enteritidis, and 'produce' for ETEC and *Norovirus* – whereas for other
266 agents, the distribution of food sources is more heterogeneous (*Campylobacter jejuni* and VTEC/STEC)
267 (figure 3b).

268 The case epidemiology of each causative agent is presented in figure 4. The two agents with a generally low
269 number of cases per outbreak, VTEC/STEC and *Salmonella* Typhimurium (figure 4a), were responsible for
270 outbreaks with high percentages of cases hospitalized (figure 4b), opposed to e.g. *Norovirus* which produced
271 outbreaks with more cases but a low percentage of hospitalization. The distribution of cases by sex (figure
272 4c), shows that ETEC, *Norovirus* and *Salmonella* Enteritidis had a higher median number of cases among
273 men, whereas *Campylobacter jejuni* and *Salmonella* Typhimurium monophasic had a higher median number
274 of cases among women. Such differences could be confounded by the setting in which outbreaks occurred
275 and a possible predominant presence of men or women, or the food sources involved and their eventual
276 higher consumption by men or women, and thus it does not necessarily indicate an association between a
277 specific sex and a causative agent. Further investigation is needed in order to determine the significance and
278 cause of the observed differences. The same consideration applies to the distribution of cases by age
279 category for each agent (figure 4d). For example, the high median number of VTEC/STEC cases among
280 young children could tend to be attributed to an outbreak caused by unpasteurized milk in a
281 'school/kindergarten' (record *VTEC_STEC_DK_2016_001* in the catalogue). However, that outbreak caused
282 a total of 5 cases in this age category, whereas an outbreak of disseminated cases caused by organic cured
283 beef sausage caused 18 cases in young children (record *VTEC_STEC_MultiCountry_2007* in the catalogue).

284 Among the most frequent causative agents, *Salmonella* Enteritidis had the highest percentage of travel-
285 related outbreak strains (figure 5a), and ETEC the highest percentage of strains associated to multiple-
286 country outbreaks (figure 5b). Outbreaks of ETEC and *Norovirus* were mostly caused by strains originating in

287 foreign countries, and in half or more of the outbreaks of *Campylobacter jejuni*, *Salmonella* Enteritidis and
288 the monophasic variant of *Salmonella* Typhimurium the country of origin could not be determined (figure 5c).
289 Determining the country of origin of outbreak strains was particularly challenging when the food source was
290 a buffet meal, a composite meal or eggs (figure 5d). Whereas eggs could justify why the origin of more than
291 half of the *Salmonella* Enteritidis strains was unknown, buffet- and composite-meals were not predominantly
292 associated with any of the three causative agents (figure 3b).

293 The Sankey diagrams in figure 6 show the distribution of *Salmonella enterica* outbreak cases for the three
294 most common serotypes, split by the three first levels of food source categorization, sequentially: food type,
295 food subtype, food subtype variety. This sort of visualization is particularly useful when data is reported at
296 different levels of detail, allowing to investigate both the distribution of records within each level individually
297 and the relationship between sequential levels. Figure 6a shows that the largest proportion of *Salmonella*
298 Enteritidis outbreaks could be identified as originating in eggs, but that the majority of the egg product
299 subtypes was either unknown or unreported. Diagrams 6b and 6c show a clear difference in the proportions
300 of meat-related outbreaks that were attributed to pork and to beef, with *Salmonella* Typhimurium outbreaks
301 being predominantly caused by pork, and its monophasic variant having a more even distribution between
302 beef- and pork-associated outbreaks.

303 The current membership policy of GFI is 'restricted', i.e. new members must request the VRE managers to
304 join. GFI membership is currently managed by the Technical University of Denmark. In order to request
305 membership, users must first create an account in D4Science⁴. Membership can then be requested at
306 <https://services.d4science.org/web/foodborneoutbreak> .

307 4. Discussion

308

309 The data catalogue has been populated with records from Danish foodborne outbreaks to better illustrate the
310 potential of GFI to new members and to encourage their contribution with their own records. We expect that
311 the perception of the catalogue's structure and of the opportunities offered by other features of the virtual
312 research environment are enhanced by providing the users with real outbreak records from a range of
313 causative agents, typing methods and food sources. However, community support is one of the main
314 challenges for the sustainability of any VRE (Candela et al., 2013). There is therefore the need for an
315 engaging dissemination strategy, including demonstration of GFI features to the target community (i.e.
316 international actors in foodborne outbreak investigation), and their motivation to contribute with data
317 reporting. The potential of GFI is expected to grow significantly with the expansion of the content of the
318 catalogue in terms of number of reported outbreaks and their diversity (e.g. geographical origin). Several
319 VREs have been successfully implemented to date in the e-infrastructure of D4Science to serve a wide
320 variety of scientific scenarios (Assante et al., 2019c). GFI is among the first implemented in the area of Food
321 and Agriculture and therefore it is expected to attract many interested users. Users from countries where
322 international network-wide communication is already practiced during outbreak investigation may show more
323 resistance to join GFI, for considering that it offers no added value. However, GFI offers the opportunity to
324 connect several presently distinct networks, into a unique, multidisciplinary (e.g. across causative agents),
325 international community. The establishment of such a collaborative community is one of the present major
326 challenges in the field of food safety.

327 An additional sustainability challenge is the attraction of further funding for the maintenance and continuous
328 adaptation of the VRE to its community's needs (Candela et al., 2013). We believe that by introducing GFI
329 and its catalogue to international research consortia studying foodborne diseases, we might identify potential
330 partners interested in populating the catalogue and eventually performing future developments in the VRE.
331 Research projects which aim at capacity building across countries in the area of food microbiology are of
332 particular interest. Future developments of GFI may include features that enable private chats and
333 conference calls among members. Such features will enhance the utility of GFI for its community, for

⁴ D4Science homepage: <https://services.d4science.org/home>

334 example by enabling straightforward, instant creation of dedicated communication channels across
335 institutes/countries during an outbreak investigation.

336 The main objective of GFI and its data catalogue is to facilitate foodborne outbreak investigations worldwide,
337 by supporting risk based sampling strategies during traceback investigations. In practice, once an outbreak
338 strain is identified and typed, it will be possible to search the data catalogue for previous outbreaks caused
339 by the same or a similar causative agent, worldwide. The user will have access to both detailed records and
340 the contact of the person(s) who reported outbreaks - both may be of special relevance to help in an ongoing
341 outbreak investigation. The result of a targeted catalogue search may indicate for example types of food
342 previously identified as an agent source, a pathogen subtype involved in a similar age/gender case
343 distribution, or the identification of the occurrence of a specific pathogen in a certain food product in a
344 country from which a suspect food product is imported. All of these inferences may help directing an
345 outbreak investigation. Once kept alive and up to date, GFI may also assist in the identification of emerging
346 food safety threats, worldwide. Overall, GFI is a platform to encourage and facilitate international
347 collaboration and data exchange between foodborne outbreak investigators.

348

349 4.1 Conclusion

350 We developed a data catalogue of foodborne outbreaks, to be shared and populated among an international
351 community of outbreak investigators. The catalogue includes for every record detailed information of
352 causative agent typing results, epidemiological data and the food source of the outbreak. It can
353 accommodate in a harmonised way most common agents and food sources involved in foodborne infections,
354 and reporting of a wide range of typing methods. It is hosted in the GFI virtual research environment, which
355 offers a shared workspace, a social networking area, and open-source software for data visualization and
356 analysis on the cloud. GFI hence represents a solution to the challenge of establishing a global collaborative
357 community of actors in foodborne outbreak investigation, which accommodates members with various
358 microbial typing practices.

359

360 5. Funding sources

361

362 This work was supported by the European Union Research and Innovation programme Horizon 2020 [grants
363 nr 643476 and nr. 731001].

364

365 6. Acknowledgements

366

367 We thank Mathias Filter and Taras Günter (Federal Institute for Risk Assessment, Germany) for their support
368 in the implementation of the VRE.

369 We thank Steen Ethelberg, Katrine Grimstrup Joensen, Mia Torpdahl and Eva Møller Nielsen (Statens
370 Serum Institute, Denmark) for their assistance in data collection and data curation.

371 We thank Sophie Honoré (Technical University of Denmark) for introducing the outbreak records into GFI.

372

373

374 7. References

375

- 376 Assante, M., Boizet, A., Candela, L., Castelli, D., Cirillo, R., Coro, G., Fernández, E., Filter, M., Frosini, L.,
 377 Kakalettris, G., Katsivelis, P., Knapen, R., Lelii, L., Lokers, R., Mangiacrapa, F., Pagano, P., Panichi,
 378 G., Penev, L., Sinibaldi, F. & Zervas, P. (2019a) Realising a Science Gateway for the Agri-food: the
 379 AGINFRA PLUS Experience. In 11th International Workshop on Science Gateways (IWSG 2019), 12-
 380 14 June 2019
- 381 Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., Lelii, L., Mangiacrapa, F., Marioli, V.,
 382 Pagano, P., Panichi, G., Perciante, C. & Sinibaldi, F. (2019b). The gCube system: Delivering Virtual
 383 Research Environments as-a-Service. *Future Generation Computer Systems*, 95, 445-453. DOI:
 384 <https://doi.org/10.1016/j.future.2018.10.035>.
- 385 Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., Lelii, L., Mangiacrapa, F., Pagano, P.,
 386 Panichi, G. & Sinibaldi, F. (2019c). Enacting Open Science by D4Science, *Future Generation*
 387 *Computer Systems*, 101, 555-563. DOI: <https://doi.org/10.1016/j.future.2019.05.063>.
- 388 Berthold, M.R., Cebon, N., Dill, F., Gabriel, T.R., Kötter, T., Meinel, T., Ohl, P., Sieb, C., Thiel, K. &
 389 Wiswedel, B. (2007). *KNIME: the Konstanz information miner*. In: Gaul W., Vichi M., Weihs C. (Eds.),
 390 *Studies in Classification, Data Analysis, and Knowledge Organization*. Springer. ISSN: 1431-8814
- 391 Candela, L., Castelli, D. & Pagano, P. (2013). Virtual Research Environments: An Overview and a Research
 392 Agenda. *Data Science Journal*, 12, GRDI75-81. DOI: <https://doi.org/10.2481/dsj.GRDI-013>
- 393 ECDC. (2019). The European Surveillance System (TESSy). Accessed at
 394 <https://ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy>
- 395 EFSA & ECDC. (2015). The European Union summary report on trends and sources of zoonoses, zoonotic
 396 agents and food-borne outbreaks in 2014. *EFSA Journal*, 13(12):4329. DOI:
 397 <https://doi.org/10.2903/j.efsa.2015.4329>
- 398 Gopinath, G., Hari, K., Jain, R., Mammel, M.K., Kothary, M.H., Franco, A.A., Grim, C.J., Jarvis, K.G.,
 399 Sathyamoorthy, V., Hu, L., Datta, A.R., Patel, I.R., Jackson, S.A., Gangiredla, J., Kotewicz, M.L.,
 400 LeClerc, J.E., Wekell, M., McCardell, B.A., Solomotis, M.D., & Tall, B.D. (2013). The Pathogen-
 401 annotated Tracking Resource Network (PATRN) system: A web-based resource to aid food safety,
 402 regulatory science, and investigations of foodborne pathogens and disease. *Food Microbiology*, 34(2),
 403 303-318. DOI: <http://dx.doi.org/10.1016/j.fm.2013.01.001>
- 404 Gormley, F. J., Little, C. L., Rawal, N., Gillespie, I. A., Lebaigue, S. & Adak, G. K. (2011). A 17-year review of
 405 foodborne outbreaks: Describing the continuing decline in England and Wales (1992-2008).
 406 *Epidemiology and Infection*, 139(5), 688–699. DOI: <https://doi.org/10.1017/S0950268810001858>
- 407 Guzman-Herrador, B., Vold, L., Berg, T., Berglund, T. M., Heier, B., Kapperud, G., Lange, H. & Nygård, K.
 408 (2016). The national web-based outbreak rapid alert system in Norway: Eight years of experience,
 409 2006-2013. *Epidemiology and Infection*, 144(1), 215–224. DOI:
 410 <https://doi.org/10.1017/S095026881500093X>
- 411 Joensen, Katrine G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., & Scheutz, F. (2015). Rapid and
 412 easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *Journal*
 413 *of Clinical Microbiology*, 53(8), 2410–2426. DOI: <https://doi.org/10.1128/JCM.00008-15>
- 414 Krause, G., Altmann, D., Faensen, D., Porten, K., Benzler, J., Pfoch, T., Ammon, A., Kramer, M.H. & Claus,
 415 H. (2007). SurvNet electronic surveillance system for infectious disease outbreaks, Germany.
 416 *Emerging Infectious Diseases*, 13(10), 1548–1555. DOI: <https://doi.org/10.3201/eid1310.070253>
- 417 Marler Clark LLP, P. (2018). Foodborne Illness Outbreak Database. Accessed at
 418 <http://www.outbreakdatabase.com>

- 419 Nadon, C. , Van Walle, I. , Gerner-Smidt, P., Campos, J. , Chinen, I. , Concepcion-Acevedo, J. , Gilpin, B. ,
420 Smith, A.M. , Kam, K.M. , Perez, E. , Trees, E. , Kubota, K. , Takkinen, J. , Nielsen, E.M., Carleton, H.
421 & FWD-NEXT Expert Panel (2017). PulseNet International: Vision for the implementation of whole
422 genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveillance*,
423 22(23):pii=30544. DOI: <https://doi.org/10.2807/1560-7917.ES.2017.22.23.30544>
- 424 Peters, T., Bertrand, S., Björkman, J.T., Brandal, L.T., Brown, D. J., Erdösi, T., Heck, M., Ibrahim, S.,
425 Johansson, K., Kornschöber, C., Kotila, S.M., Le Hello, S., Lienemann, T., Mattheus, W., Nielsen,
426 E.M., Ragimbeau, C., Rumore, J., Sabol, A., Torpdahl, M., Trees, E., Tuohy, A. & De Pinna, E. (2017).
427 Multi-laboratory validation study of multilocus variable- number tandem repeat analysis (MLVA) for
428 *Salmonella enterica* serovar Enteritidis, 2015. *Euro Surveillance*, 22(9):pii=30477. DOI:
429 <https://doi.org/10.2807/1560-7917.ES.2017.22.9.30477>
- 430 R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical
431 Computing, Vienna, Austria. <https://www.R-project.org/>
- 432 Richardson, L.C., Bazaco, M.C., Parker, C.C., Dewey-Mattia, D., Golden, N., Jones, K., Klontz, K., Travis,
433 C., Kufel, J.Z. & Cole, D. (2017). An Updated Scheme for Categorizing Foods Implicated in Foodborne
434 Disease Outbreaks: A Tri-Agency Collaboration. *Foodborne Pathogens and Disease*, 14(12), 701-710.
435 DOI: <https://doi.org/10.1089/fpd.2017.2324>
- 436 RIVM. (2018). Databases. Accessed at <https://www.rivm.nl/en/noronet/databases>
- 437 RStudio Team (2016). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA.
438 <http://www.rstudio.com/>
- 439 Timme R.E., Sanchez Leon M. & Allard M.W. (2019). *Utilizing the Public GenomeTrakr Database for*
440 *Foodborne Pathogen Traceback*. In: Bridier A. (Eds) *Foodborne Bacterial Pathogens.*
441 *Methods in Molecular Biology*, vol. 1918, pp 201-212. New York: Humana. DOI:
442 https://doi.org/10.1007/978-1-4939-9000-9_17
- 443 Wielinga P.R. ,Hendriksen, R.S., Aarestrup, F.M., Lund, O., Smits, S.L., Koopmans, M.P.G. & Schlund, J.
444 (2017). *Global Microbial Identifier*. In: Deng X., den Bakker H., Hendriksen R. (Eds), *Applied*
445 *Genomics of Foodborne Pathogens. Food Microbiology and Food Safety*, pp 13-31. Springer, Cham.
446 DOI: https://doi.org/10.1007/978-3-319-43751-4_2
- 447 World Health Organization (WHO) (2015). *WHO estimates of the global burden of foodborne diseases:*
448 *foodborne disease burden epidemiology reference group 2007-2015*. Geneva: World Health
449 Organization. ISBN: 9789241565165
- 450 Anon. (2006). Annual Report on Zoonoses in Denmark 2005. The Ministry of Family and Consumer Affairs,
451 Copenhagen, Denmark. ISSN: 0909-4172.
- 452 Anon. (2012). Annual Report on Zoonoses in Denmark 2011. National Food Institute, Technical University of
453 Denmark. ISSN: 0909-3837.
- 454 Anon. (2013). Annual Report on Zoonoses in Denmark 2012. National Food Institute, Technical University of
455 Denmark. ISSN: 1600-3837.
- 456 Anon. (2017). Annual Report on Zoonoses in Denmark 2016. National Food Institute, Technical University of
457 Denmark. ISSN: 1600-3837.
- 458
- 459

Figure captions

Figure 1: Features of the GFI virtual research environment

Snapshots of the virtual research environment: *a* - welcome page; *b* – data catalogue; *c* –shared workspace; *d* – data analytics options and R studio; *e* – social networking

Figure 2: Typing methods used in foodborne outbreaks of *Salmonella enterica* subsp. *enterica* over time

Percentage of *Salmonella enterica* subsp. *enterica* strains typed by MLVA, PFGE, Phagotyping and WGS by year, between 2005 and 2016, in Denmark.

Figure 3: Outbreak settings and major food sources of most frequent causative agents

Relative distribution of outbreak settings (*a*) and food sources (*b*) for the most common causative agents of foodborne outbreaks occurred in Denmark, between 2005 and 2016.

Figure 4: Epidemiology of outbreaks caused by the most frequent causative agents

Distribution of the number of cases per outbreak (*a*), percentage of cases hospitalized per outbreak (*b*), number of cases by sex (*c*) and number of cases by age class (*d*) for the most common causative agents of foodborne outbreaks occurred in Denmark, between 2005 and 2016.

Figure 5: Country of origin of most frequent causative agents and food sources

Relative distribution of travel- and domestic- related outbreaks (*a*) and of multiple- and single- country outbreaks (*b*), and relative distribution by country of origin (*c*) of the most common causative agents of foodborne outbreaks occurred in Denmark, between 2005 and 2016. Relative distribution by country of origin of the food sources of foodborne outbreaks occurred in Denmark, between 2005 and 2016 (*d*).

Figure 6: Distribution of food sources of three *Salmonella enterica* subsp. *enterica* serotypes

Sankey diagrams showing the distribution of food sources for foodborne outbreaks occurred in Denmark, between 2005 and 2016, by *S. Enteritidis* (*a*), *S. Typhimurium* (*b*) and *S. Typhimurium* monophasic (*c*).

Table captions

Table 1: Description of fields in the GFI data catalogue.

Table 2: Number of outbreaks, outbreak strains and cases per year available in the catalogue at time of public launch.

Table 3: Number of records by causative agent and typing method available in the catalogue at time of public launch.

Table 4: Number of records by food category available in the catalogue at time of public launch.

Journal Pre-proof

Table 1

Field	Field description
Title	Title of the outbreak record (agent_country_year)
Description	Short description of the outbreak (agent, country, year, food source)
Tags	Record tags (agent, country, year, food source,...)
Data and Resources	Scientific publication, report and/or data related to outbreak
Item URL	URL of the outbreak record
Causative agent	
CausativeAgent	Causative agent of the outbreak
AMRProfile	AMR profile of the outbreak strain
AMRMethod	AMR method used
Genogroup	Norovirus genogroup of the outbreak strain
GenogroupProtocol	Norovirus genogroup protocol used
MLST	MLST type of the outbreak strain
MLSTScheme	MLST scheme used
MLVAProfile	MLVA-type of the outbreak strain
MLVAProtocol	MLVA protocol used
PFGETyped	Was the outbreak strain PFGE typed?
PFGEProtocol	PFGE protocol used
Phagetype	Phagetype of the outbreak strain
PhagetypeStandard	Phagetype standard used
Serotype	Serotype of the outbreak strain
SerotypeMethod	Serotyping method used
VirulenceProfile	Virulence profile determined for the outbreak strain
VirulenceProfileDetectionMethod	Method used to determine virulence profile
WasTheIsolateSequenced	Was the outbreak strain Whole Genome Sequenced (WGS)?
OtherTypingInformation	Additional typing information, incl. WGS accession number, if available
Epidata	
Cases	Number of patients involved in the outbreak
Countries	Country/countries involved in the outbreak
Hospitalized	Number of patients involved in the outbreak who were hospitalized
Men	Number of male patients involved in the outbreak
MultipleStrains	Was more than one outbreak strain involved in the outbreak? (Preferably, cases should be reported per outbreak strain. If cases cannot be attributed to individual strains, they should be reported in different records for each strain)
Old	Number of patients above 65 years old involved in the outbreak
OutbreakDuration	Duration of the outbreak in months (outbreaks lasting one month or less should be reported as 1 month)
ReferenceOtherStrain	Title of the record(s) referring to other strain(s) involved in the same outbreak
Setting	Specific setting for the outbreak (EFSA and ECDC, 2015)
SingleCountryOutbreak	Was there only one country involved in the outbreak?
StartMonth	Month for first registered case in outbreak
StartYear	Year for first registered case in outbreak
TravelRelated	Was the outbreak travel related? (Travel = travel abroad in a seven-day period prior to disease onset)

Field	Field description
	(Anon., 2017)
TravelCountry	If the outbreak was travel related, to which country have the cases travelled to?
UnknownGender	Number of patients with unknown gender involved in the outbreak
Women	Number of female patients involved in the outbreak
Young	Number of patients below 5 years old involved in the outbreak
<i>Food source</i>	
MajorFoodType	Food sources were categorized based on a hierarchic food categorization scheme from the Interagency Food Safety Analytics Collaboration (IFSAC) (Richardson et al., 2017; see appendix...).
MajorFoodSubtype	
FoodSubtypeVariety	
A_ProcessingPreparationConsumption	
B_ProcessingPreparationConsumption	
CountryOriginOfContamination	Country where food was contaminated with the disease causing agent
FoodOriginOfContamination	Production level where food was contaminated with the disease causing agent
OtherFoodsourceInfo	Specific food source causing the outbreak (e.g. raspberries), or other additional information related to the foodsource
<i>Report details</i>	
EvidenceForOutbreak	Evidence for outbreak (epidemiological and/or microbiological)
Reference	URL(s) to relevant scientific literature for the outbreak
<i>Management Info</i>	
Author	Name of user who reported the outbreak
Maintainer	Name of user who maintains the record
Version	Record version
State	State of the record (e.g. active)
Last Updated	Date of last record update
Created	Date for record creation

AMR - antimicrobial resistance

Table 2

Year	Outbreaks	Strains	Cases
2005	10	10	1263
2006	5	6	533
2007	4	23	429
2008	7	13	1825
2009	6	6	786
2010	25	54	1091
2011	5	5	4024
2012	9	10	546
2013	8	14	729
2014	7	7	525
2015	6	6	296
2016	10	12	1419
Total	102	166	13466

Table 3

Causative agent		Serotype	PFGE type	Phage type	MLVA profile	MLST	AMR profile	Virulence profile	Genogroup	WGS
Campylobacter	3									1
<i>Campylobacter jejuni</i>	6					2	1			3
<i>Clostridium perfringens</i>	2									
<i>Cryptosporidium hominis</i>	1									1
Enteropathogenic <i>Escherichia coli</i> (EPEC)	3	3						1		
Enterotoxigenic <i>Escherichia coli</i> (ETEC)	22	21								
Shiga toxin/verocytotoxin-producing <i>Escherichia coli</i> (VTEC/STEC)	8	6	1			3		5		3
Other pathogenic <i>Escherichia coli</i>	2	0					1	1		
<i>Entamoeba histolytica</i>	1									
<i>Giardia intestinalis/lambliaduodenalis</i>	1									
Hepatitis A virus	2									
<i>Listeria monocytogenes</i>	4	1	2			2				2
Norovirus	42								27	22
Other parasitic agents	3									
Rotavirus	2									
<i>Salmonella enterica</i> subsp. enterica	54	54	7	25	37	1	30			15
Sapovirus	1									
<i>Shigella sonnei</i>	4		1				1			
<i>Staphylococcus aureus</i>	1									
Streptococcus	1		1							
<i>Yersinia enterocolitica</i>	3	1								
Total	166	86	12	25	37	8	33	7	27	47

Table 4

MajorFoodType	MajorFoodSubType	FoodSubTypeVariety	A_ProcessingPreparati onConsumption	B_ProcessingPreparati onConsumption						
Fish	1									
Dairy	1	Fluid milk	1	Unpasteurized 1						
Meat-Poultry	29	Meat	25	Beef	8	Ready-to-eat 6	Acidified/Fermented 1			
				Pork	16			Salt-cured 1		
								Fully-cooked 4		
								Raw/other processed 1		
		Poultry	3	Chicken 2		Non-intact raw 3	Ground beef 2			
		Other poultry	1							
Eggs	6	Shell Eggs	3							
Produce	38	Fruits	9	Small	9	Frozen 9				
				Vegetables	29			Sprouts	2	
								Seeded	3	Fresh 25
								Row crops	22	
Herbs	1									
Composite meal	3									
Buffet meal	10									
Beverages	5									
Unknown	9									
Total	102		70		64		45	8		

The table presents only category levels for which there are presently records available in the catalogue. For a complete list of all levels available for reporting see Appendix 1.16.

a

Global Foodsource Identifier Administration Members Social Networking Wiki Analytics Data Token Calendar

Journal Pre-proof

Welcome to the Global Foodsource Identifier Virtual Research Environment!

Global Foodsource Identifier (GFI) is a restricted-access virtual research environment designed for the global community that investigates **foodborne outbreaks**. Its aim is to provide a comprehensive and harmonised framework for the international sharing of microbial typing results from foodborne outbreaks, and to facilitate the collaboration and interaction between outbreak investigators, globally. The systematic data sharing by GFI members leads to the compilation of an extensive catalogue of outbreak data including detailed typing information of the causative agent, epidemiological data and the identified food source for every record. Such data may support future outbreak investigations among the GFI community by potentially suggesting suspect food items at an early stage, according to the identification of the same pathogen type in past outbreaks.

GFI includes a searchable **data catalogue** which is maintained by the contribution of the community members. The catalogue can be searched for records of past outbreaks where a specific foodborne pathogen type (e.g. serotype, PFGE, MLVA, MLST, etc) has been identified, a specific country has been affected, or a specific setting or food source has been involved.

GFI also offers the possibilities to:

- interact in real-time through a **social networking area**
- import the catalogue data into Rstudio or Jupyter Notebook for **analysis**
- relevant documentation and the results of data analysis can be disseminated to other users in a **shared workspace**

This VRE was originally conceived within the research project **COMPARE**, which represents a multidisciplinary effort to enable global data sharing and analytics for the rapid identification, containment and mitigation of emerging infectious diseases and foodborne outbreaks.

b

Items About

Campylobacter

24 items found for "Campylobacter" Order by: Relevance

OWNER Campylobacter_jenuni_DK_2006
Outbreak of Campylobacter jenuni in Denmark in 2006.

OWNER Campylobacter_jenuni_DK_2016_000
Outbreak of Campylobacter Jenuni in Denmark in 2016 due to duck meat.

OWNER Campylobacter_jenuni_DK_2016_001
Outbreak of Campylobacter Jenuni in Denmark in 2016.

OWNER Campylobacter_DK_2015
Outbreak of Campylobacter in Denmark in 2015.

OWNER Campylobacter_jenuni_DK_2005
Outbreak of Campylobacter jenuni in Denmark in 2005 due to raw chicken.

Followers 1 Items 24

Filter by location

d

Causative agent	Serotype Method	Serotype			
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Bareilly	no	-	-
Other parasitic agents	-	NA	no	-	-
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Typhimurium	yes	Standard PulseNet Salmonella PFGE	Anderson
Clostridium perfringens	-	NA	no	-	-
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Typhimurium	no	-	Anderson
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Typhimurium	no	-	Anderson
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Typhimurium	no	-	-
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Typhimurium	no	-	-
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Typhimurium	no	-	-
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Typhimurium	yes	Standard PulseNet Salmonella PFGE	Anderson
Salmonella enterica subsp. enterica	White-Kauffmann-Le Minor scheme (Grimont and Weil, ...)	Typhimurium	no	-	Anderson

c

Workspace > VRE Folders > FoodborneOutbreak

COMPARE's workspace

- VRE Folders
- DataMiner

File actions: New Folder, Upload, Download, Refresh, Delete, Rename

Files:

- Shared documents
- Wiki files
- R studio files
- Jupyter files
- Catalogue data exported
- Data analysis - shared results
- ToU - GFI.pdf
- GetCatalogueData5

e

Members Social Networking Wiki Analytics Data Token Calendar

Share updates

Share an update or a link, use "@" to mention and "#" to add a topic

Notify members: OFF ON

Share

News feed

Show sorted by: newest Post

COMPARE Compare just now

The virtual research environment Global Foodsource Identifier (GFI) is open!

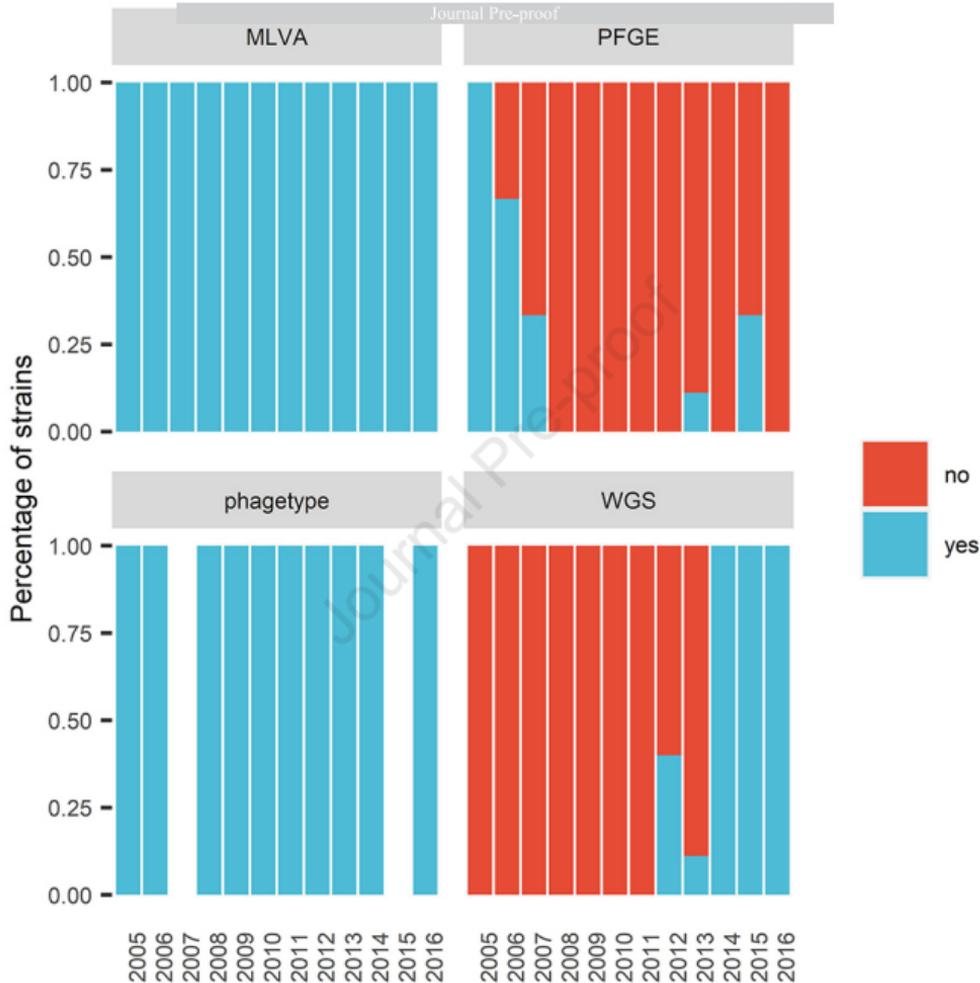
GFI offers its community:

- a shared catalogue of foodborne outbreak records including multiple pathogens and a comprehensive set of typing methods and protocols, where members can search among existing data and share their own records
- additional features that allow document exchange, communication and data visualization and analysis between members

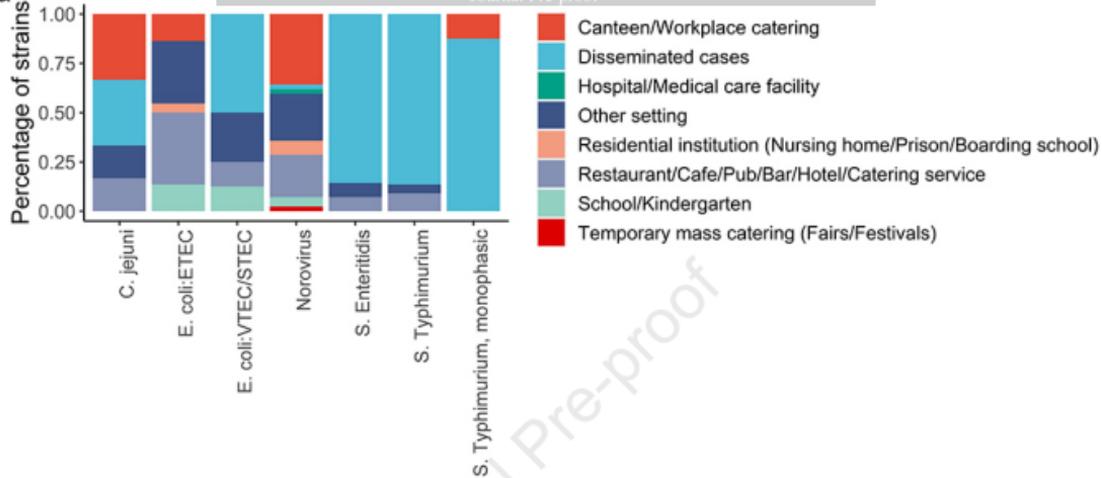
Community engagement is key for the sustainability of a collaborative environment, and the potential of GFI will continuously grow with the expansion of data sharing and collaboration between its members.

We look forward to welcoming you as an active member.

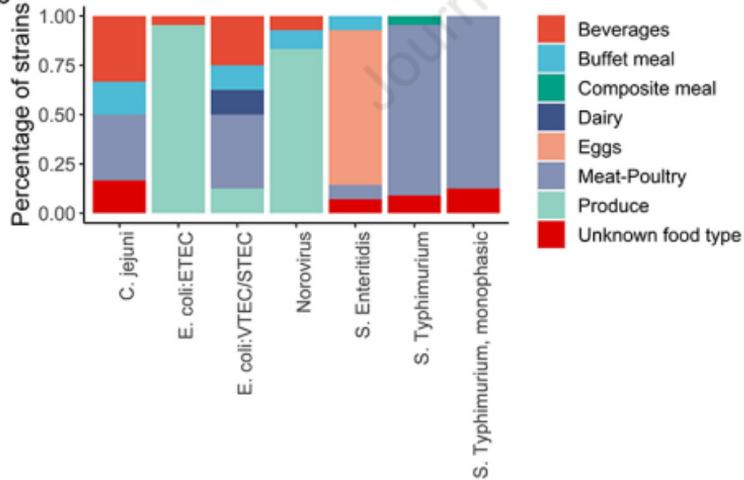
Reply Like

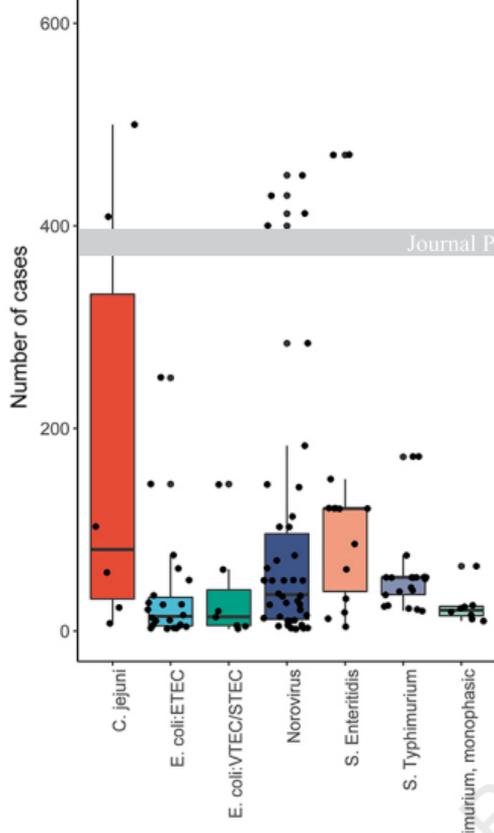
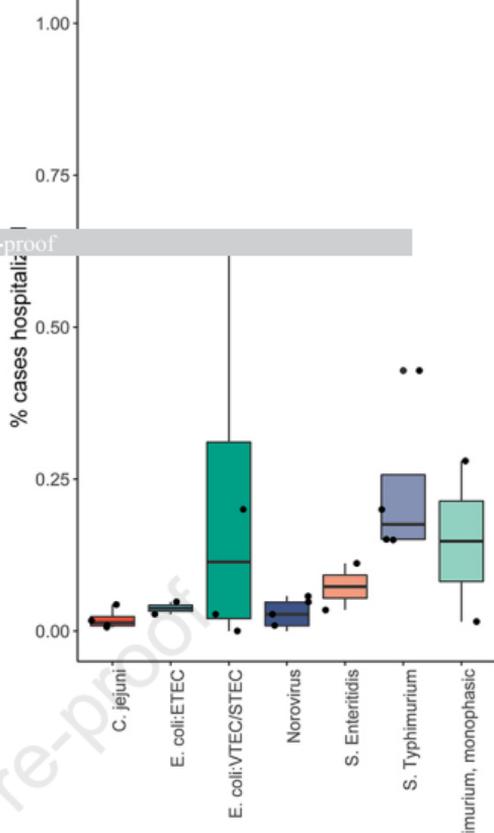
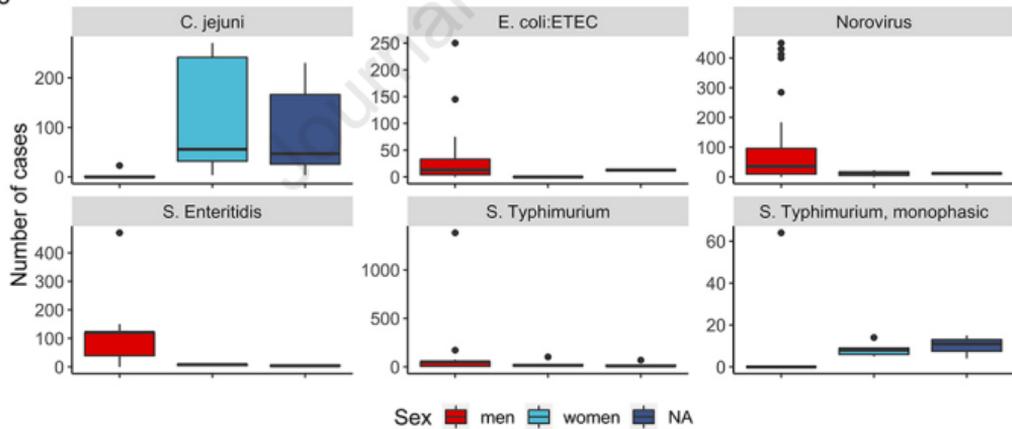
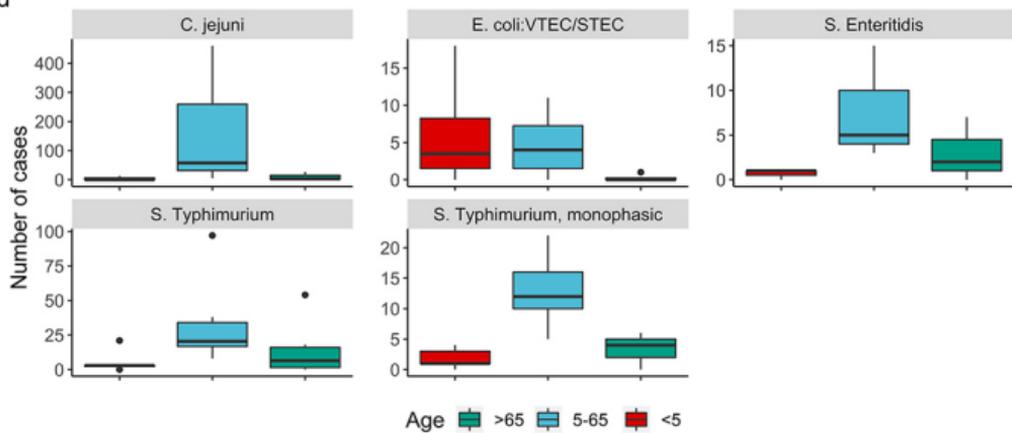


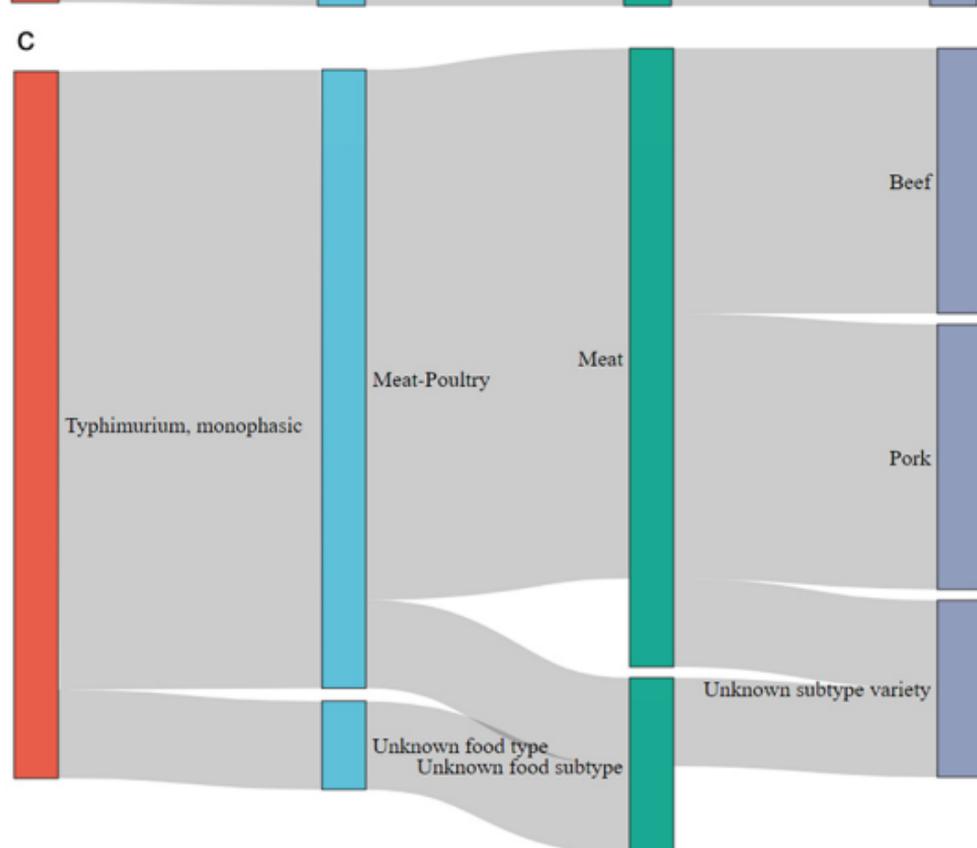
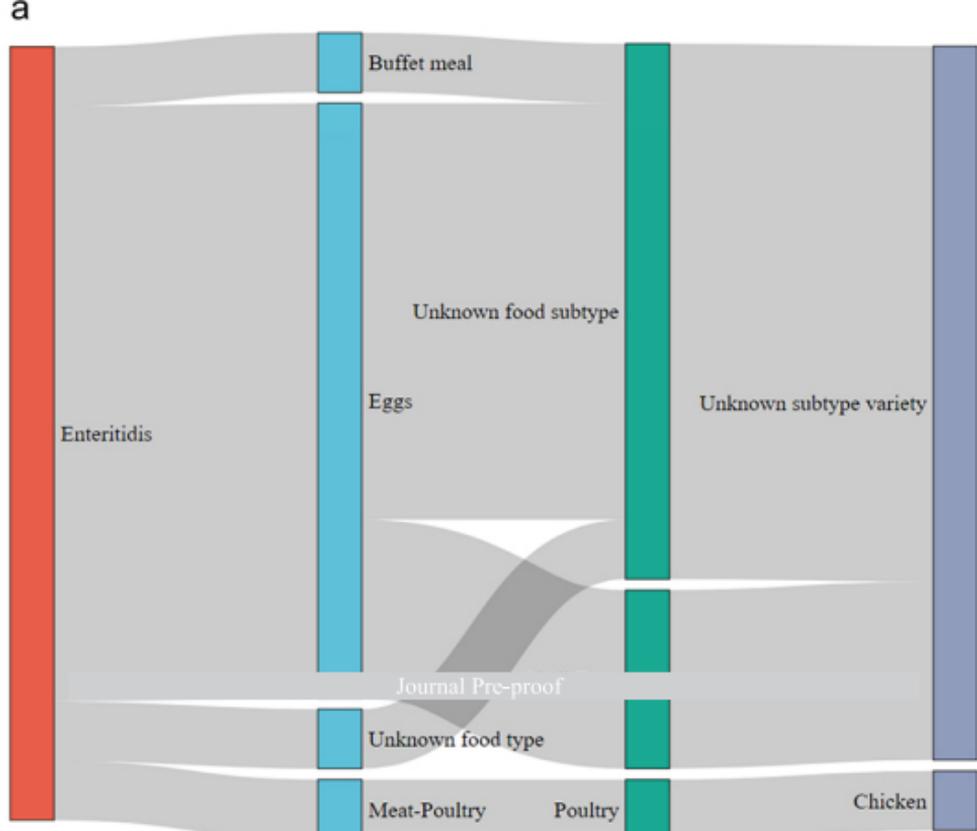
a



b



a**b****c****d**



Global Food-source Identifier (GFI): collaborative virtual research environment and shared data catalogue for the foodborne outbreak investigation international community

Ana Sofia Ribeiro Duarte^{1*}, Cecilie Liv Nielsen^{1,4}, Leonardo Candela², Lars Valentin³, Frank M. Aarestrup¹, Håkan Vigre¹

¹Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark. Kemitorvet 204, DK 2800 Kongens Lyngby, Denmark

²Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", National Research Council of Italy, Via G. Moruzzi, 1 - 56124 Pisa, Italy

³ Unit Food Hygiene and Technologies, Supply Chains, Food Defense, Department Biological Safety, German Federal Institute for Risk Assessment (BfR), Max-Dohrn-Strasse 8-10, 10589, Berlin, Germany

⁴Department of Animal Science – ANIS Welfare, Aarhus University. Blichers Alle 20, DK 8830 Tjele, Denmark.

*corresponding author - contacts: asrduarte@gmail.com; Kemitorvet 204, DK 2800 Kongens Lyngby, Denmark

Highlights

- Global Food-source Identifier is a public data catalogue of foodborne outbreaks.
- GFI exists in a virtual research environment that facilitates global collaboration.
- GFI supports a community with various microbial typing practices.
- GFI may provide evidence to guide risk-based sampling in outbreak investigations.

Journal Pre-proof

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof